Automatic and Instructed Attention in Learned Predictiveness

Lauren T. Shone (lsho0771@uni.sydney.edu.au)
Evan J. Livesey (evan.livesey@sydney.edu.au)
School of Psychology, Brennan McCallum Building,
University of Sydney, NSW 2006 AUS

Abstract

In novel situations, learning is biased towards information that has a degree of prior predictive utility. In human learning, this is termed the learned predictiveness effect and has proved critical in theorising about the role of attention in learning. Two experiments are reported in which the relative contribution of controlled and automatic processes to learned predictiveness are investigated. Experiment 1 showed that while learned predictiveness is susceptible to instructional manipulation, this effect is partial. Experiment 2 manipulated predictive utility and instruction orthogonally in order to test the potential involvement of automatic processes. It was found that even when cues were explicitly instructed as causal, learning was biased in favour of previously predictive over previously non-predictive cues. Interestingly, this was reversed for cues instructed as irrelevant. This suggests that learned predictiveness benefits attentional control, whereby information is both easier to attend and ignore.

Keywords: human learning, attention, controlled processing, automatic processing

Introduction

An important question facing theories of associative learning is the nature of the relationship between learning and attention. Accordingly, many associative theories (e.g., Kruschke, 2001; Mackintosh, 1975; Pearce & Hall, 1980) accept that stimulus selection is influenced by attentional processes. Such theories share the basic assumption that the attention devoted to a stimulus is flexible, and governed by its past utility in predicting events. Importantly, this will subsequently influence the rate at which a stimulus enters into future associations.

Evidence in favour of learned attention originates from experiments in which past predictive utility biases learning in a novel situation. A robust example, first reported by Le Pelley and McLaren (2003; see also Lochmann & Wills, 2003), is the learned predictiveness effect. The basic experimental design used to demonstrate the effect is shown in Table 1. Participants are initially exposed to a scenario in which they are required to learn a causal relationship between cues and outcomes. Each trial consists of the presentation of a compound of two cues, leading to one of two outcomes. Critically, each compound consists of one perfectly predictive cue (represented by A – D), and one non-predictive cue (W – Z). For example, A is consistently paired with the outcome O1, and therefore has perfect predictive utility. Alternatively, W has no predictive utility because it is paired equally often with both outcomes O1 and O2.

Once these relationships have been learned, a novel scenario is introduced. The same cues, in novel combinations, are then employed in order to predict different outcomes. Importantly, although the cues are again presented in compound, this time neither component has superior predictive utility. That is, both A and W are perfect predictors as they share the same objective relationship with outcomes O3 and O4 respectively. What differs between the components of the new compounds is their status as a predictive or non-predictive cue in the initial stage of learning. Subsequent tests reveal that more is learned about the relationship between previously predictive cues and the new outcomes compared to previously non-predictive cues.

Table 1. *A typical learned predictiveness design*.

Phase 1	Phase 2	Test
AW – O1	AY – O3	AD
AX – O1	BZ – O4	XY
BW – O2	CW – O4	BC
BX – O2	DX – O3	WZ
CY – O1		
CZ – O1		
DY – O2		
DY – O2		

Note. Letters indicate individual cues. O1 – O4 refer to four outcomes.

Traditionally, this bias, consistently replicated across various scenarios (see Le Pelley, 2010, for a recent review), has been interpreted to suggest that attention is modulated by the difference in predictive validity during initial stages of learning. According to this logic, attention to A – D will be high following phase 1 and will therefore have an advantage when entering into new associations during the second phase. This effect has proved critical in theorising about the reciprocal nature of the relationship between human learning and attention.

The learned predictiveness effect is consistent with models of associative learning that assume attention changes according to mechanisms of associative competition (e.g., Mackintosh, 1975; Le Pelley, 2004; Pearce & Mackintosh, 2010). For example, Mackintosh (1975) proposed that changes in the association between a cue and an outcome are governed by both attention paid to the cue and the discrepancy between the occurrence of the outcome and the extent to which it is already predicted on the basis of that cue, that is, the prediction error for an individual cue. Critically, attention to the cue changes according to a

comparison between its prediction error and the prediction error for other cues available at the same time. The cues with smaller individual prediction errors (i.e. those with higher predictive utility) will command more attention as learning proceeds. Higher attention, in turn, drives faster learning.

Despite its replicability, the exact nature of the learned predictiveness effect has only recently been questioned. Indeed, the concept of attention is associated with a variety of cognitive mechanisms (see Pashler, 1998; Wright & Ward, 2008, for a review), raising the question of which processes critically characterise the effect. For example, in demonstrations of learned predictiveness there is often a high degree of conceptual similarity between scenarios. One possibility, therefore, is that the effect is governed by a simple heuristic arising from inferential reasoning. That is, it is possible that participants make the explicit assumption that the predictive utility of cues A – D will transfer across similar contexts (Mitchell, Griffiths, Seetoo, and Lovibond, 2012).

According to this explanation, learned predictiveness should be susceptible to manipulations of inferred beliefs. Indeed, Mitchell, et al., (2012) have provided evidence in support of this view. In their Experiment 2, inferences were directly manipulated across phases by way of instruction. At the onset of the second phase, participants in the continuity condition were explicitly instructed that the same cues would be relevant. Alternatively, those in the change condition were instructed the opposite, that previously predictive cues were now irrelevant. Critically, this condition revealed a complete reversal of the effect. That is, more was learned about the relationship between previously irrelevant cues and the novel outcomes. That learned predictiveness is sensitive to variations in explicit reasoning suggests a role for controlled, volitional attentional processes in explaining the effect.

However, there is evidence to suggest that the presence of the inference alone is not sufficient to produce the learned predictiveness effect. For example, Le Pelley et al. (2010a) investigated the expression of learned predictiveness adopting a procedure in which the critical relationships were embedded in text form. Interestingly, they failed to observe the effect; the attentional bias was only observed when the relevant information was presented in trial and error form across multiple trials. This is contrary to what would be expected if explicit causal attribution was the sole mechanism responsible for this bias. Similarly, related paradigms have found opposing influences of training and instruction on learned attentional responses (Le Pelley, Mitchell, & Johnson, 2013). Taken together these findings raise the possibility that learned predictiveness reflects the operation of a combination of inferential and non-inferential processes.

As noted previously, learned predictiveness has taken an important role in theorising about learned attention. A common feature of such theories is the assumption that attentional changes are automatic in response to the formation of associations between events (e.g., Kruschke, 2001; Le Pelley, 2004; Mackintosh, 1975; Pearce & Mackintosh, 2010). According to this view, because associations between predictive cues and outcomes increase rapidly during phase 1 of a learned predictiveness experiment, these cues are automatically attended. Thus, previously predictive cues will capture attention at the start of phase 2, such that associations between these cues and novel outcomes are facilitated. Importantly, this process does not rely on a deliberate attempt by the individual to control attention in a biased fashion according to the nature of the phase 1 relationships.

While the results of Mitchell et al. (2012) appear to oppose this explanation, there is reason to suggest that their experimental design did not provide the conditions under which the presence of automatic processes could be adequately detected. For example, their demonstration relies on a definitive manipulation: Non-predictive cues were explicitly emphasised as important. If it is assumed that controlled attention is capable of modulating the expression of automatic processes, given the appropriate conditions, then it is possible that the manipulation was too strong, overriding the influence of automatic attention. Thus, although this manipulation demonstrates that learned predictiveness is susceptible to voluntary control via instruction, it does not test whether automatic processes also contribute to the effect under uninstructed conditions.

Further, the scenario employed, in which fictitious seeds grow different trees, potentially favours a more categorical inferential process whereby the outcome is most likely attributable to only one of the cues and not the other. This aspect of the design may have facilitated a complete reversal based on conceptual aspects of the scenario in addition to the manipulation of interest.

Therefore, the relative contribution of controlled and automatic processes to the learned predictiveness effect remains to be fully specified. The aim of the present experiments was to investigate this relationship.

Experiment 1

Experiment 1 made use of the same instructional manipulation employed by Mitchell et al. (2012), albeit with a different cover scenario, in order to replicate their original result. The allergist scenario, employed in numerous demonstrations of learned predictiveness (e.g., Le Pelley & McLaren, 2003) was used in which participants were asked to play the role of a doctor who must discover the allergies of a fictitious patient. The cues consisted of different foods, which predict the occurrence of various allergic reactions, serving as outcomes. At the start of phase 2, a new patient was introduced who consumed the same foods, but suffered novel reactions. As before, participants were required to discover which foods were leading to which reactions. The structure of the training phases is shown in Table 1 and reflects the standard learned predictiveness design. At the start of phase 2, one group of participants (the "same"

condition) were told that it was likely that both patients were allergic to the same foods, whereas those in the "change" condition were instructed that their two patients likely suffered from allergies to different foods.

In line with the findings of Mitchell et al. (2012), we anticipated that the bias in learning observed in learned predictiveness would proceed according to the instructions issued at the start of phase 2 training.

Method

Participants Forty-eight University of Sydney students (27 female, 21 male; age 18 – 24) participated in the experiment.

Apparatus and Stimuli All experiments were conducted on Apple Mac Mini computers attached to a 17-in. monitor, and programmed in PsychToolbox for Matlab (Brainard, 1997; Pelli, 1997). Foods were randomly allocated for each participant to serve as cues A – Z in the experimental design, and consisted of: Coffee, Fish, Lemon, Cheese, Eggs, Garlic, Bread, and Peanuts. Similarly, four allergic reactions were randomly allocated to serve as the four outcomes, and were: Headache, Nausea, Rash, and Fever.

Procedure After being randomly allocated to either the same or change conditions, participants were instructed that their task was to learn which foods were causing which allergic reactions in a fictitious patient. They were told that on every trial, two foods that the patient had eaten would be presented. On being shown the foods, participants were required to predict which of two allergic reactions would occur.

Phase 1 consisted of the eight trial types shown in Table 1. Each of these was presented once in each of 16 blocks of trials. The order of trials was randomised across blocks. Each trial was followed by feedback stating whether their prediction was correct, as well as providing the actual allergic reaction experienced.

At the start of phase 2, participants were told that they now had a new patient and, as before, would be required to learn which foods were causing which allergic reactions. Those in the same condition were told that their new patient was allergic to the same foods as their previous patient, whereas those in the change condition were instructed that their new patient was allergic to different foods.

Phase 2 consisted of 16 blocks, each of which contained one of the four trial types shown in Table 1. As before, trial order was randomised within blocks and feedback was provided after each trial.

A test phase was administered immediately following phase 2. All cues were presented individually and in a randomised order throughout this phase. On each test trial, a cue would appear and participants were asked to indicate whether the cue had been paired with outcome 3 or outcome 4. This was done by making a rating on a linear analogue scale, labelled "*Definitely goes with [outcome 3]"* on the left anchor, and "*Definitely goes with [outcome 4]"* on the right anchor.

Finally, a manipulation check was included to ensure that participants had remembered the instructions at the start of phase 2. Participants were presented with both sets of instructions and required to report which of those applied to their patient. There were no exclusions on the basis of this check.

Results

Phase 1 For each block, accuracy was averaged across the eight compound trials to gauge acquisition. Accuracy increased consistently across training. A mixed-measures analysis of variance (ANOVA) with block (1-16) and condition (same vs. change) as factors revealed a significant main effect of block, $F(15, 690) = 40.1$, $p < .001$, but no significant effect of condition, $F < 1$, and no block × group interaction, $F < 1$, suggesting that the two groups learned at an equivalent rate in phase 1.

Phase 2 A mixed-measures ANOVA examining phase 2 acquisition showed a significant effect of block, $F(15, 690) = 44.13$, $p = < .001$, as well as a significant block × group interaction, $F(15, 690) = 2.1$, $p < .05$. The effect of condition did not reach significance, $F(1, 46) = 3.95$, $p = .053$.

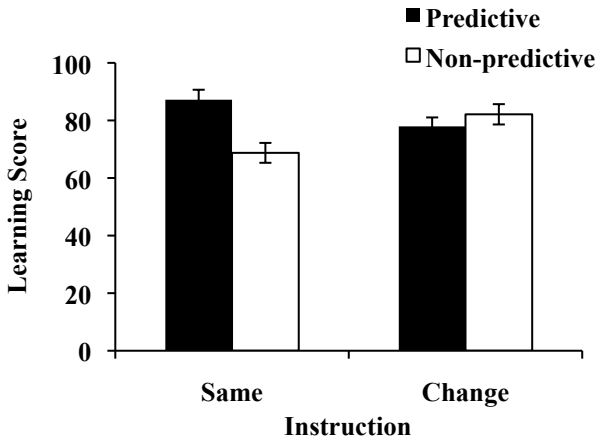

Figure 1. Learning scores for the same and change conditions for previously predictive and previously non-predictive cues

Test data A learning score for each cue was calculated by combining accuracy for memory of the cue-outcome pairings in the test phase with the magnitude of the rating. This yielded a score out of 100 for each cue, with higher scores indicating better retention. Scores could range between 100 and -100. Scores were averaged according to whether they were predictive (A – D) or non-predictive (W – Z) in phase 1. These are shown for the same and change conditions in Figure 1.

Scores were subjected to a mixed-measures ANOVA with group (same vs. change) and cues (predictive vs. non-predictive) as factors. Averaged over cue, there was no significant difference between the same and change conditions, $F < 1$. Similarly, there was no effect of cue, $F < 1$. However, as suggested by Figure 1, this resulted from a significant cue × group interaction, $F(1, 46) = 8.79$, $p < .05$.

This was further investigated with a simple effects analysis, which revealed that learning scores for predictive cues was higher than non-predictive cues in the same condition, $F(1, 23) = 11.51$, $p < .05$. The difference between predictive and non-predictive cues did not differ significantly in the change condition, $F < 1$.

Discussion

Our data provide a partial replication of Mitchell et al. (2012). While the same condition showed a standard learned predictiveness effect, this was abolished rather than reversed in the change condition. That is, there was no difference between previously predictive and previously non-predictive cues when participants were told that non-predictive cues were informative for the second phase.

Overall, a clear effect of instruction was observed making use of a scenario in which it is less likely that causal attribution is biased towards categorical reasoning. This suggests that the result of Mitchell et al. (2012) is not entirely a consequence of the conceptual structure of their scenario, further validating the influence of voluntary control on learned predictiveness.

However, it is important to note that our reversal was incomplete in the critical condition. On the basis of the current design, it is unclear why this should be the case. It is possible that the results from the change condition reflect competition between opposing inferential and automatic processes. While automatic processes would bias learning in favour of previously relevant cues, explicit inference favours irrelevant cues.

Alternatively, there may be added difficulty in the change condition. If more is learnt about the predictive cues in phase 1, this means that they may be required in order to confirm the new object of attention, that is, the previously irrelevant cue. That is, if the explicit identity of the previously irrelevant cues is uncertain due to the fact that little learning has proceeded to these cues, then previously relevant cues may be actively used to guide responding. This is an additional process that is not necessary in the same condition.

Given that the reversal design does not allow the contribution of automatic processes to be assessed, Experiment 2 used an orthogonal manipulation of predictiveness in phase 1 and instruction to further test the relative contribution of voluntary and automatic processes.

Experiment 2

Experiment 1 confirmed that learned predictiveness is susceptible to the manipulation of inferred beliefs. In Experiment 2, we aimed to further test the involvement of automatic processes. This was done by orthogonally manipulating the predictive status of cues in the first phase and the instructional manipulation. The design of Experiment 2 is shown in Table 2. The first phase of training was identical to that seen in Experiment 1. At the end of the initial training phase, all participants were told explicitly which foods the new patient was allergic to. However, two of those cues were previously predictive, while two were previously non-predictive. That is, they were told that the new patient was allergic to cues A and C, and X and Z.

This means that there were two cues (A and C) that were predictive in phase 1, and known to cause allergies in the new patient, and two previously predictive cues (B and D) known not to be allergens. Similarly, of the previously non-predictive cues, two (Z and X) were now known to cause allergies, and the remaining two (Y and W) known to be safe. The design therefore creates the condition in which an unambiguous instructional manipulation is present without removing the opportunity to observe an automatic influence of phase 1 training, if indeed it is present.

Table 2. *Design of Experiment 2.*

Phase 1	Phase 2	Test
AW – O1	**A**Y – O3	A
AX – O1	B**Z** – O4	B
BW – O2	**C**W – O5	C
BX – O2	D**X** – O6	D
CY – O1		W
CZ – O1		X
DY – O2		Y
DY – O2		Z

Note. Letters indicate individual cues. Underlined letters indicate cues instructed as informative for phase 2. O1 – O6 refer to six outcomes.

If, as suggested by the findings in Experiment 1, controlled processes are in operation, then a clear influence of instruction should be observed whereby more will be learned about cues A, C, X, and Z in the second phase. However, if automatic attention favouring predictive cues is also present, then a difference should also be observed between instructed cues according to whether they were relevant (A and C) or irrelevant (X and Z) in the first phase. Given the advantage conferred by predictive utility, this predicts that more should be learned about A and C compared to X and Z.

Method

Participants Participants comprised twenty-four University of Sydney students (20 female, 4 male; age 18 – 23).

Apparatus and Stimuli Experimental stimuli remained the same as that employed in Experiment 1, with the exception that two additional allergic reactions were introduced to account for added outcomes in the design. These were Coughing and Sweating.

Procedure Phase 1 training and instructions remained identical to that used in Experiment 1. Following phase 1, participants were told that they were now observing the allergies of a new patient, but that they would be provided with a set of foods that the patient was allergic to. They were shown the names of four foods, corresponding to cues A, C, X, and Z and were informed that they would need to learn which of these corresponded to the various reactions that the patient was experiencing.

Given that foods were named explicitly, a shorter phase 2 with fewer trials per cue was employed. Participants completed four blocks, each block consisting of one of the four trial types shown in Table 2. On each trial, participants were now required to predict which of four allergic reactions would occur.

During test, each cue was displayed individually in random order. The four outcomes were displayed on screen and participants were asked to indicate which of these the cue had been paired with. This was followed by the appearance of a rating scale, asking how confident they were in their response. The left anchor was labelled "*Not at all confident"*, and the right anchor labelled "*Very confident"*.

Finally, the manipulation check required participants to report the instructed allergens of the second patient. Five participants were excluded, having failed to report this content, leaving 19 participants in the analysis.

Results

Phase 1 Acquisition across blocks increased steadily for phase 1. A repeated-measures ANOVA showed a significant main effect of block on accuracy, $F(15, 270) = 13.01$, $p < .01$.

Phase 2 Overall, accuracy increased during phase 2, resulting in a significant main effect of block on accuracy, $F(3, 54) = 12.95$, $p < .01$. However, acquisition varied according to whether a compound contained an instructed component that was previously predictive or an instructed component that was previously non-predictive, such that accuracy was significantly higher for the former (AY/CW higher than BZ/DX), $F(1, 18) = 7.25$, $p < .05$. The interaction was not significant, $F < 1$.

Test data Accuracy scores, shown in Figure 2, were subjected to a repeated-measures ANOVA with predictiveness (predictive vs. non-predictive) and instruction (instructed vs. ignored) as factors. This revealed a significant main effect of instruction, $F(1, 18) = 18.28$, $p < .01$, as well as a significant instruction × predictiveness interaction, $F(1, 18) = 10.6$, $p < .01$. The effect of predictiveness failed to reach significance, $F < 1$.

A simple effects analysis investigating the interaction showed that for instructed cues, accuracy was significantly higher for previously predictive cues, $F(1, 18) = 5.7$, $p < .05$. Interestingly, this was reversed for the remaining cues, such that accuracy was significantly higher for previously non-predictive cues, $F(1, 18) = 6.4$, $p < .05$.

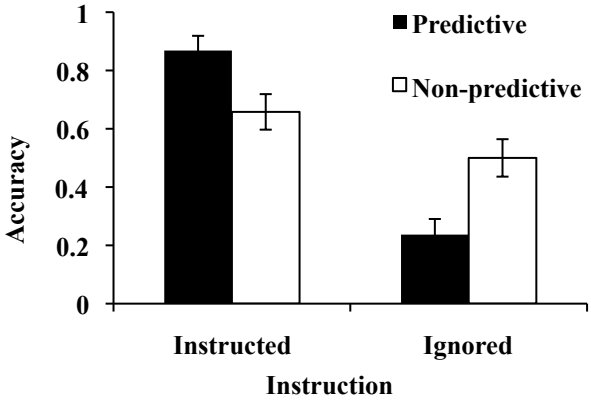

Figure 2. Accuracy scores for previously predictive and previously non-predictive cues at test in Experiment 2 for the instructed and ignored conditions.

Discussion

Consistent with the findings in Experiment 1, there was clearly an effect of instructional manipulation. However, the learned predictiveness effect was still evident amongst cues known to be allergenic. That is, more was learned about the previously predictive cues compared to previously non-predictive cues, despite the explicit knowledge that both sets of cues were allergens. This is consistent with the involvement of automatic processes transferred from initial learning.

However, it is interesting to note that the opposite pattern emerged for cues that were not instructed as allergens, and would presumably be ignored by participants. Thus it appears that previously predictive cues were easier to ignore when known to be irrelevant. This may reflect a general benefit of prior predictive utility whereby attention is more easily directed either towards or away from stimuli in novel situations.

Alternatively, the difference in acquisition during phase 2 between compounds that contained instructed components that were previously predictive (AY and CW) and non-predictive (BZ and DX) raises the possibility that some sort of automatic interference from phase 1 means that less is learned in general about phase 2 compounds in which participants have to attend to the previously non-predictive cue and ignore the previously predictive cue. If these

compounds were indeed more difficult to learn, despite explicit instruction, this would result in the observed lower accuracy for instructed, yet previously irrelevant cues at test.

General Discussion

The experiments reported above suggest that a purely inferential account of learned predictiveness is insufficient to fully characterise the effect. However, it is clear that proposing an additive influence of inferential reasoning and automaticity is similarly inadequate as the results reported here suggest an interaction between the two.

For example, in phase 2 of Experiment 2, participants were given information that directly informed them which cues the patient was and was not allergic to. Even though participants could have ignored the non-causal cues completely, some learning of the cue-outcome relationships was evident. The result of interest regarding these non-causal cues was that previously predictive stimuli were learned about more poorly than previously nonpredictive stimuli. If the effects of the prior predictive history of the cues simply added or subtracted from selective attention in an automatic fashion then one would expect the opposite result for this incidental learning. That is, the predictive cues should be learned about more readily than the nonpredictive. This result suggests an interaction between control of attention and the effects of prior predictive history, which is not explained by either an inferential account nor the conventional associative account of learned predictiveness.

Accordingly, there are a growing number of studies that show that the learned predictiveness effect does not operate via the competitive associative algorithms of attentional change described by Mackintosh (1975; Le Pelley, 2004; Pearce & Mackintosh, 2010). For instance, Le Pelley et al., (2010b) found that competition between cues in compound was not necessary for learned predictiveness to occur, and Livesey et al. (2011) found no evidence that *direct* comparison between predictive and nonpredictive cues affected the magnitude of learned predictiveness at all. The current study demonstrates another way in which the automatic allocation of attention appears to behave differently from model predictions. Although there appears to be a relatively automatic influence of the previous history of the cues, that influence only matches the predictions of associative learning theories for cues that are deliberately attended and not those that are deliberately ignored.

Clearly an important step in implementing attentional processes within models of human learning will require further investigations into the mechanisms responsible for biases in learning related to past predictive utility. Such biases remain to be fully specified with regards to how information is attended and ignored.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45, 812–863.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology, 57(B),* 193 – 243.

Le Pelley, M. E. (2010). Attention and human associative learning. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 187–215). Oxford: Oxford University Press.

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *Quarterly Journal of Experimental Psychology, 56*B, 68–79.

Le Pelley, M. E., & Mitchell, C. J., & Johnson, A. M. (2013). Outcome Value Influences Attentional Biases in Human Associative Learning: Dissociable Effects of Training and Instruction. *Journal of Experimental Psychology: Animal Behavior Processes,* 39, 39 – 55.

Le Pelley, M. E., Reimers, S. J., Calvini, G., Spears, R., Beesley, T., & Murphy, R. A. (2010a). Stereotype formation: Biased by association. *Journal of Experimental Psychology: General, 139,* 138–161.

Le Pelley, M. E., Turnbull, M. N., Reimers, S. J., Knipe, R. L., & Murphy, R. A. (2010b). Learned predictiveness effects following single-cue training in humans. *Learning & Behavior, 38,* 126–144.

Livesey, E. J., Thowart, A., De Fina, N. L., & Harris, J. A. (2011). Comparing Learned Predictiveness Effects Within and Across Compound Discriminations. *Journal of Experimental Psychology: Animal Behavior Processes,* 37, 446 – 465.

Lochmann, T., & Wills, A. J. (2003). Predictive history in an allergy prediction task Proceedings of EuroCogSci 03: *The European Conference of the Cognitive Science Society* (pp. 217–222). Mahwah, NJ: Erlbaum.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement.*Psychological Review, 82,* 276–298.

Mitchell, C. J., Griffiths, O., Seetoo, J., & Lovibond, P. F. (2012). Attentional mechanisms in learned predictiveness. *Journal of Experimental Psychology: Animal Behavior Processes, 38,* 191–202.

Pashler, H. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Wright, R. D., & Ward, L. M. (2008). *Orienting of attention.* New York: Oxford University Press.